```
Set     Items   Description
S1         10   (METACRAWLER? OR METASEARCH?)(5N)DUPLICAT?(2N)(FILTER? OR -
             REMOV?)
S2          5   RD (unique items)
S3          5   S2 NOT PY>2001
File 202:Info. Sci. & Tech. Abs. 1966-2003/Jul 31
         (c) 2003, EBSCO Publishing
File  47:Gale Group Magazine DB(TM) 1959-2003/Aug 11
         (c) 2003 The Gale group
File 484:Periodical Abs Plustext 1986-2003/Aug W3
         (c) 2003 ProQuest
File 553:Wilson Bus. Abs. FullText 1982-2003/Jul
         (c) 2003 The HW Wilson Co
File  15:ABI/Inform(R) 1971-2003/Aug 20
         (c) 2003 ProQuest Info&Learning
File  13:BAMP 2003/Aug W1
         (c) 2003 Resp. DB Svcs.
File 148:Gale Group Trade & Industry DB 1976-2003/Aug 19
         (c)2003 The Gale Group
```

3/9/2 (Item 1 from file: 47)
DIALOG(R)File 47:Gale Group Magazine DB(TM)

05161014 SUPPLIER NUMBER: 20330335 (THIS IS THE FULL TEXT)

Toward more comprehensive Web searching: single searching versus megasearching.

Notess, Greg R.

Online, v22, n2, p73(4)
March-April, 1998
ISSN: 0146-5422 LANGUAGE: English RECORD TYPE: Fulltext; Abstract
WORD COUNT: 2719 LINE COUNT: 00214

ABSTRACT: Searches of the leading Web search engines often provide different results, raising the question to how best to conduct comprehensive Web searches. Two basic approaches are available: searching each searching engine individually and using a megasearch engines. The disadvantages and advantages of each approach are discussed. Unfortunately, megasearch engines such as Dogpile, Inference Find, and MetaFind have limitations.

TEXT:

In my October DATABASE column, I explored the size and overlap of the major Web indexes and discovered that these search engines have much less overlap than might be expected. Instead, each database has unique records. Given the number of unique items in each database and the lack of duplication, we should think about the best approach for finding information using the major search engines.

Using a multiple engine search tool, such as Inference Find, Dogpile, and MetaFind, seems like one appropriate response. Unfortunately, these also have significant limitations in their processing of search syntax and built-in limits on the number of hits retrieved from each database. A second approach is to search the largest of the databases one by one, using the command language unique to each to increase the precision of the search. Both of these methods are useful, but neither fully addresses the problem of trying to run a comprehensive search of Web sites. What are the advantages and disadvantages to each approach?

SINGLE SEARCH TOOL APPROACH

While the lack of completeness in databases of Web sites is a problem, that certainly does not mean that these are small databases. With page counts ranging from 30 to 50 million, general searches bring up huge numbers of hits. Intelligent application of phrase, Boolean, and field searching can be used quite effectively to more precisely narrow search results.

Using a single search engine has the definite advantage of being able to exploit all of that search engine's best features, Infoseek's follow-up search, truncation in AltaVista, page depth limits on HotBot, and Excite's sort options are all examples of important search features available when using a single search tool.

Note that the intertwined nature of the Web offers a wide variety of ways to connect to each search engine. Not everyone who searches Excite goes directly to www.excite.com. The search buttons on Netscape Navigator and Internet Explorer both take the user to special miniature versions of selected search engines. These sites are visited and used by large numbers of people each day. In addition, the search engines themselves offer advice on adding HTML code in your pages for embedding a miniature search form or creating specialized search boxes on the browsers. Multiple search engines and desktop search engines can also send out queries.

Yet, the fullest range of features, the most up-to-date options, and specialized databases are only available by going directly to the search engine itself. The companies change their search interfaces and add new features frequently. In most cases, these changes are not reflected immediately in the other programs and Web sites that send queries to that same database.

AltaVista is a good example. It has a simple and advanced search mode. Both support a variety of field searches and truncation of up to five characters with *. The advanced search includes a date limit, and it recently added a language limit. None of these features can be used effectively when AltaVista is searched with a megasearch engine or as a follow-up to a Yahoo! search. HotBot's many search options pose a similar challenge. To use the advanced search features of a specific search engine, going directly to that database's Web site is going to be the best way of searching.

Problems with the Single Tool Approach

The two primary negatives of using a single search tool are the learning curve and the incompleteness of the databases. Using just one of the search engines has the advantage of only having one search syntax to learn, just as getting a single brand of CD-ROMs requires learning only a single search system. But, due to the lack of overlap, the most effective searchers will still be those familiar with all the search features of the major search engines.

The incompleteness of the large databases of Web sites is a more significant problem. Understanding all the most advanced features of HotBot cannot help find a record that is not in their database. With the single tool approach, the only answer is to run successive searches on each of the largest databases. For a comprehensive search, be sure to search HotBot, AltaVista, Excite, and Infoseek. The necessity for sequential searching is what makes the megasearch engines look so attractive.

THE MEGASEARCH APPROACH

Megasearch engines, also known as parallel search engines or multiple search engines, use one form that simultaneously sends a single query to a number of search engines and then presents the results. The advantage of using these tools is the parallel processing of the search, with each search engine running at the same time. In addition, the better ones present a variety of options for sorting the results and **duplicate removal** .

Online megasearch engines include **MetaCrawler** , Inference Find, Dogpile, and MetaFind. Desktop megasearch engines run on a local computer but still send out parallel search requests and sort results. These include WebCompass, WebSeeker, and EchoSearch.

In addition to the speed that results from the parallel processing, megasearch engines can gather some of the unique records in each of the databases. Their capabilities for sorting results by host, keyword, date, or search engine can make a long list of results much easier to browse and more informative. Instead of presenting just ten results at a time, most megasearch engines list more hits on a single page, which also makes browsing the answer set easier.

Problems with Megasearchers

But beware. The parallel search processing gives the impression of solving the unique record problem inherent in the single search approach. By searching all the largest search engines, the end result should be the most comprehensive search available. Unfortunately, since few people other than librarians ever want a comprehensive search, the reality is that most megasearch engines' results are far from comprehensive.

The typical approach of a megasearch engine is to grab the top 10-50 hits from each search engine and then combine those with the results from the other databases. Sometimes the user can customize the number of hits retrieved from each search engine, but that rarely exceeds 50 per database. A phrase search for unemployment rate on MetaCrawler finds 82 hits when the number of hits from each search engine is set to the maximum of 30. However, a search on Excite (one of the search engines used by MetaCrawler) finds more than 43,000 hits for the same phrase.

Other problems with megasearch engines include faulty handling of Boolean searches, the inability to use field searching or other search enginespecific advanced features, and the absence of some of the largest search engines. Most of the multiple search engines do not include Northern Light, and many exclude HotBot, even though those have two of the largest Web databases. Especially with the desktop megasearch engines, the results will sometimes include extraneous links--advertisements and help links that are on the search engine's results page, but which have no relation to the search request.

Inference Find, Dogpile, and MetaFind are three megasearch engines that offer some advantages for comprehensive searching, even while suffering from some of the problems, and we will explore each in a little more detail.

INFERENCE FIND

An example of one of the better megasearch engines is Inference Find (http://www.inference.com/ifind/). The search tool not only sends out the search statements at the same time, but also merges the results, removes duplicates, organizes the hits into logical groups, and lists hits on a single line.

Inference Find uses only six databases: WebCrawler, Yahoo!, Lycos, AltaVista, Infoseek, and Excite. The free version of Inference Find available on the Internet cannot be customized but the commercial version designed for use on an intranet can be modified to point to different search engines.

Inference Find does not do any processing of the search statement to match it to the specific search engine. Thus, using quotes to designate a phrase will work in all the databases except Lycos (unless it is modified to point to Lycos Pro). HotBot is not included as one of the databases, so no hits will be pulled from HotBot.

In terms of comprehensiveness, the Inference Find documentation states that it "retrieves the maximum number of results each engine will allow. Infoseek, for example, is one of those that will return only 10 items at a time. To get the most out of it, InfoSeek is called three times in parallel to retrieve 30 items." Try a search for a common word on Inference Find to see how far short it falls from the "maximum" In two examples I tried, both came back with fewer than 150 hits all together. In addition, Inference Find does not show which results came from which search engine. However, for those searches that can be expected to retrieve fewer than 100 hits, Inference Find can be an effective and quick tool for finding those sites.

DOGPILE

Dogpile and MetaFind are two other megasearch engines that go at least part way toward overcoming the defects common to many multiple search tools. Despite its name, Dogpile is an impressive searching tool (http://www.dogpile.com). Created to search the smaller subject directories first and then the larger Web databases, it includes 14 Web databases, five Usenet news databases, three FTP lists, and three news databases. Not only does Dogpile do an excellent job of correctly handling Boolean and phrase searching, it also displays the syntax of how the search was submitted to each search engine. This can be of great assistance to help searchers decide for themselves whether a follow-up search direct to a specific database is necessary.

Unlike other parallel search engines, Dogpile does not send the search phrase exactly as typed. Its documentation details a translation process, stating that Dogpile takes the query and "processes it so that you will get the maximum benefit from your search." Searchers can use the NEAR operator and Dogpile will substitute AND for those engines that do not support NEAR. Phrase searches can be designated with quotes, and Dogpile will remove them for search engines that do not support their use.

Dogpile's default method of searching the smaller, targeted subject directories first and then the larger automated Web databases is a sound strategy. For broad topics or general queries, it should work fairly well.

However, Dogpile also allows the user to customize and choose exactly which databases are searched. This ability to customize can be used effectively to search across all five of the largest search engines: HotBot, AltaVista, Excite, Infoseek, and Lycos. To do this, choose the Customize Dogpile option. Put those five search engines in the top five slots and mark all the other slots as Skip. Dogpile will then save these settings in a cookie on your computer and offer a URL (http://www.dogpile.com/custom/) for you to bookmark that presents a form for searching your chosen databases.

Like other multiple engine search tools, Dogpile still limits retrieval to the first ten records from each of the chosen databases. Thus, this technique works best for rare or unique words or phrases. To its credit, Dogpile search results include links to the next set of documents from each of the databases. It also reports the total number of hits from each database at the beginning of the listing. The first three databases to respond are displayed first. The user then needs to request the results from the next set of three. This can become a tedious process for browsing results.

Dogpile's customization capability and detailed reporting on how the search statements are structured make it a useful megasearch engine, especially for translating a single query to match the requirements of specific search engines. While the display limit of ten hits per database and the lack of a sort option make it difficult to browse, the link into the full results sets from the individual databases means that Dogpile can be used to try to approach a comprehensive search, even if it takes clicking on quite a few links.

Dogpile should also be commended for making an important and truthful statement in its documentation that applies equally well to other megasearch engines: "Search engines change their format all the time. Thus (Dogpile) is guaranteed not to work 100% of the time with 100% of the engines."

METAFIND

For another approach to multiple searching, the makers of Dogpile offer MetaFind (http://www.metafind.com). Rather than starting with the subject directories, MetaFind searches through just seven databases and combines the results into a single display. Like Dogpile, MetaFind limits the number of hits it retrieves from each database. The search engines used are as follows, with the number of hits retrieved from each in parentheses: AltaVista (20), Excite (20), HotBot (50), Infoseek (25), OpenText (10), WebCrawler (50), and PlanetSearch (30). These limits mean that it cannot be used for a comprehensive search. However, for a search that will yield only a few or no hits, it can be very effective.

One of MetaFind's most useful features is its sorting capability. These include the default sort--by title keywords--as well as unsorted, alphabetical by title, or a domain sort. The keyword sort can be a bit confusing at first. It is simply named keyword and the words listed are those in the search plus a section for "other." The actual sorting under the keyword occurs if the record has the keyword in the title. Those records where none of the search terms appears in the title are listed under the "other" section. The domain sort can be very helpful in many searches when looking for hits from a particular country or a government site. MetaFind's alphabetical title sort is a welcome relief from the typical relevance ranking sort.

While MetaFind is limited in the total number of hits retrieved, it does provide a quick way of browsing the top few results from the four largest search engines. Its integration of all the hits into one set and its very useful sort options make it a useful multiple search engine for searches that will result in relatively few hits.

SO WHAT'S A SEARCHER TO DO?

There is no perfect strategy for searching the Internet. Plenty of files and Web pages are completely hidden from the best of the Web search engines. Even if those files are unearthed, the tools now available have significant limits. However, a few general guidelines are suggested by the above discussion.

Dogpile presents a good starting point for a general search, especially for searches that should try a directory first. A customized Dogpile search can be used to run concurrently on the major databases, but it works best with searches that have a very low volume of hits.

MetaFind and Inference Find, like some of the desktop megasearch products, do a good job of integrating and sorting search results, but they do not retrieve all of the available records.

The single search tool approach is the method of choice for searches that require the use of advanced search features available from only certain search engines. A succession of single search engine queries remains the most effective way to try a comprehensive Web search. But, while we await the development of even larger databases with more advanced sorting, combination, and limit features, these tools can at least help find some information gems amidst the mounds of the mundane.

Communications to the author should be addressed to Greg R. Notess, MSU Library, P.O. Box 173320, Bozeman, MT 59717-3320; 406/1994-6563; align@montana.edu; http://imt.net/~notess/.

SPECIAL FEATURES: other; illustration
DESCRIPTORS: Online searching--Analysis; Internet/Web search services--Analysis
PRODUCT/INDUSTRY NAMES: 4811525 (Online Search Services & Directories)
SIC CODES: 4822 Telegraph & other communications
FILE SEGMENT: TI File 148

3/9/3   (Item 1 from file: 484)
DIALOG(R)File 484:Periodical Abs Plustext

04458693   (THIS IS THE FULLTEXT)

Publicity through better Web site design

Guenther, Kim

Computers in Libraries (ICLB), v19 n8, p62-67, p.5
Sep 1999
ISSN: 1041-7915   JOURNAL CODE: ICLB
DOCUMENT TYPE: Feature
LANGUAGE: English   RECORD TYPE: Fulltext; Abstract
WORD COUNT: 3254

ABSTRACT: Publicity is the key to helping users find one's web site. Tips to help one publicize their web site are given, such as listing the site with the most popular search engines.

TEXT:

Headnote:

"The question you should be asking now is, 'How will users find my site?"'

(Photograph Omitted)

Publicizing a Web site is the final stage, albeit an ongoing stage, of the Web development process. In this phase, content and design of the Web site are complete, at least for the initial launch, and the user audience you initially identified prior to developing the site should again become the focus. The question you should be asking now is, "How will users find my site?"

Publicity Is the Key

As easy as publicizing a Web site sounds, it is in fact probably one of the most confusing aspects of Web development. Unfortunately it is an area where many Web developers spend the least amount of time and thought, choosing instead to integrate the latest Web development bells and whistles rather than incorporate well-structured content and behind-the-scenes metadata. Although much of this confusion is due to lack of indexing standards among search sites, Web developers can adopt their own Web page development standards and strategies to ensure that their sites are properly indexed across different search sites. This strategy takes into account proper content organization, e.g., the pyramid writing principle; an understanding of search site indexing; use of metatag information; and appropriate subject area knowledge to properly assess how and where publicity efforts have the most impact.

The word "publicize" is defined as "to draw public attention to." In a library setting, publicity occurs at different levels depending on the type of library-special, academic, or public. Special libraries often receive publicity as part of the larger organization's marketing efforts; public libraries receive publicity from local schools and community-related publications and groups; and academic libraries are publicized by virtue of the schools, departments, and students they support. Although they benefit from the publicity wave generated by the overall marketing efforts occurring on their behalf, libraries and their respective departments can and should market their individual Web sites to their specific audiences. With this in mind, there are many ways, both traditional and nontraditional, to successfully publicize a library Web site and announce to the world, "We are here!"

Web Search Engine Sites

One of the most effective ways to publicize a Web site is to list it with the most popular search engines, Web directories, and guides. Collectively, these information locators are referred to as "search sites," and each indexes and categorizes information differently. If you've ever tried locating information on the Web you've probably used one of these search sites. Although searching the Web can be frustrating given its sheer volume, a search site helps make the task of hunting for information manageable. Over the past 5 years several search sites have risen to the top as the most comprehensive and user friendly within this highly competitive market. Yahoo!, Britannica Internet Guide, Galaxy, Lycos, AltaVista, HotBot, Excite, and Infoseek each provide a huge database of indexed Web sites updated daily and available to users for free.

Most search sites and their back-end databases work in similar fashion: A user submits a query, usually a keyword or word phrase, through the site's search page interface; the keyword or word string is checked against the site's keyword indexes; and the most relevant documents are returned as "hits" or hyperlinked entries. Relevancy is generally established by the number of times the keyword appears within the document. The more the keyword appears, the more the document is weighted with regard to its relevancy as compared to other similar documents. The end result is a list of hits arranged in descending order of relevancy. The top-most documents are considered more relevant compared to documents appearing further down the list.

Understanding how queries are processed and how relevancy is established are important for those seeking information, but they are even more important for Web site developers who can directly affect how their Web sites and Web documents are indexed by the different search sites. Understanding the subtle indexing differences between search sites will help to integrate the very best Web development practices in order to publicize a Web site effectively.

Search Engines

Search engines use indexing software programmed to travel the Internet in search of new or updated Web sites and their associated pages. Sounding more like an entomologist's watch list, indexing programs, or agents-also referred to as "spiders," "robots," "ants," "worms," or "crawlers"-go from page to page following internal and external links from the page. Essentially their programmed task is to visit every Web page on the Internet and download indexing information used to describe the page. Indexing varies among search engines and their associated spidering programs. Some gather information contained within the <HEAD> </HEAD> HTML tags, others index words found in the introductory text or collect every word contained in the document, still others extract a limited combination of all of these. The end result is an enormous database searchable by keyword and comprehensive in scope but lacking the necessary human intervention to effectively balance recall (number of hits) and relevance (quality). Adoption of standardsHTML and metadata use-that would facilitate automated indexing is still just a Web developer's dream.

Search Engine Hybrids

Search engines are considered hybrids when they also offer their own limited associated Web directory. Unlike comprehensive Web directories like Yahoo! where users submit their URLs for review, search engine hybrids tend to develop their directories in house with the input of reviewers who come across interesting sites that are then added. These sites are reviewed and rated and made available to users in a "best of the Web" or "top sites" area.

The most popular search engines and search engine hybrids are as follows:

Lycos
http://www.lycos.com (hybrid)
Excite
http://www.excite.com (hybrid)
Alta Vista
http://www.altavista.com
Infoseek
http://www.infoseek.com (hybrid)
HotBot
http:/www.hotbot.com
WebCrawler
http://www.webcrawler.com (hybrid)
Northern Light
http://www.northernlight.com

Internet Directories

Directories tend to be more accurate than search engines since they are compiled and maintained by humans. Directories facilitate serendipitous browsing, much like browsing the shelves in a library, where the user finds other related resources shelved virtually within the same category. Unlike search engines crawling the Web in search of new sites, directories require users to submit their Web addresses along with general descriptive information. This information is used to review the site, and then catalog the site if it is accepted. Reviews are often conducted by a subject specialist, who determines the best placement of the site within the directory's overall framework of subject categories. Unlike search engines where well-placed HTML tags and content structure determine better indexing results, directories do not rely on page structure except in instances where it may influence the site's review. The following are the most popular directories:

The Mining Company
http://www.miningco.com

Beaucoup
http://www.beaucoup.com

Yahoo!
http://www.yahoo.com

Internet Guides

Guides are similar to directories but are often subject based, e.g., health or business. Users submit their Web addresses along with general descriptive information, and the site is either accepted or declined following a review. Like directories, placement of HTML tags and content structure makes no difference in how the site is indexed except where it may or may not influence the review. The most popular guides are these:

The Argus Clearinghouse
http://www.clearinghouse.net

Galaxy
http://galaxy.einet.net/galaxy.html

Encyclopedia Britannica's Internet Guide
http://www.ebig.com

Metasearch Engines

Where search engines offer a product composed of an indexing agent, a database, and a searching tool, metasearch engines offer more service than product. Their services are simple: A metasearch engine queries a user's search request across several search engines simultaneously and pulls between 10 and 50 hits from each, **removing** the **duplicates** . **Metasearch** engines don't index Web sites or create and maintain their own databases but instead facilitate the search process by configuring the user's search request in a format recognized by other search engines. The most popular metasearch engines are the following:

MetaCrawler
http://www.metacrawler.com

Netlocator
http://www.netlocator.com

SavvySearch
http://www.savvysearch.com

Dogpile
http://www.dogpile.com

MetaFind
http://www.metafind.com

Adopting Your Own Web Development Standards

Lack of indexing standards places the burden of indexing a site squarely on the shoulders of the Web developer, who must write well-structured content and use appropriate HTML and metadata in order to accommodate the indexing methods of most search sites. Adopting Web development standards for design, HTML use, metadata, and content structure (writing) is an important part of the overall Web development strategy (see Figure 1). Enforcement of these standards among those developing the library's Web pages is often achieved through the creation of a Web development style guide or a formalized set of policies and procedures to ensure consistency across the site. The larger the Web site and the more decentralized its development, the more important these development standards become.

(Table Omitted)

Captioned as: Figure 1:

One of most important standards to incorporate into Web development is the integration of information contained within the header portion of the document. Although a lack of indexing standards persists among search engines, one constant among most of them is the use of metatags for indexing. Metatags are the most important pieces of information you can provide to have your site and pages indexed properly and to reap the publicity you deserve. Most search engines provide detailed information about how they incorporate the use of metatag information, including the number of characters or words that can be used in the keyword or description metatag. Metatag use does not guarantee your page will make top billing within a search engine, but it does increase relevancy for your page.

Using Metatags to Index Your Site Effectively

Metatags provide information about the document-meta-information or metadata-in an HTML document provided in the document header within the <HEAD> </HEAD> tags. The following metatags are most often used by search engines for indexing purposes: <META NAME="keywords" CONTENT="keyword I, keyword 2, keyword3, etc.">

The <META> tag keyword attribute provides words or word phrases for search engine indexing. Often these keywords are indexed along with other keywords found in the document itself. To make the most of your keywords, don't repeat the words used in your opening text. Instead use related words or word combinations. Limit your keyword list to 1,000 characters and do not repeat keywords more than seven times. "Spamindexing," that is, repeating a keyword within the <META> tag keyword attribute, is not a good strategy for increasing your site ranking. Most of the larger search engines now counter the spamindexing strategy by simply ignoring the entire tag when a keyword is repeated more than seven times. See the example below:

<META NAME="description" CONTENT="This document describes how to publicize a library Web site">

The <META> tag description attribute provides a description of the Web page for search engine indexing. Try to keep your description concise with the most important information at the beginning since many of the search engines read only the first 200 to 250 characters.

Although keyword and description represent the most often used metatags for Web site indexing, there are many other metatags that can be used to further define a Web page, such as author, document type, copyright, etc. Metatags can also support the maintenance of Web pages by redirecting browsers from expired pages using the <META> tag "expired," redirecting users to related pages using the <META> tag "refresh," and setting restrictions for automatic indexing or "spidering" of your Web site with the <META> tag "robots." You can find further information on both search engine differences and metatag use at Search Engine Watch (http:fl/www.searchengine watch.com) or The Web Developer's Virtual Library (http://www.stars.com).

Web Site Addresses and Naming Conventions

Site Web address or URL: Request a shortened Web address or "alias" providing more potential for name recognition among site users. An alias can be set up by your system administrator or by your Internet Service Provider. Here's an example of a Web address with an associated alias: Longer version of Web address: http:// www.med.virginia.edu/hs-library/info_ serv/erd

Alias: http://hsc.virginia.edu/e-reference

File names: Choose a file name that reflects the content of the document. Some search engines do retrieve file names as part of the content that is indexed. As a general rule, use lowercase letters only for file names; this aids both the user and the maintainer of the page. For example, use the file name "reference.html" instead of "Ref.html."

Page titles: Choose a title that reflects the content of the document and place it between the <IE> and <LITTLE> tags.

Writing for the Web

Some search sites automatically or manually extract information and keywords from a Web page's initial content or any of the text appearing between the <BODY> </BODY> tags. Structuring the writing of content from general to specific is called the inverted pyramid writing method and facilitates the reader's ability to read content to the desired level of specificity. Behind the scenes, the pyramid writing method accommodates search sites using initial page content and keywords for indexing. As part of a search return following a query, initial page content is often used as a very short descriptor or abstract along with the site's title, relevance ranking, date of last modification, file size, and Web address. Structuring your content using this top-down method benefits both the user of your site and the searcher.

Mass Submission Method

Another method to actively index a site is to submit Web site information to a service that will submit this information en masse to most of the popular search sites. Below is a list of the most popular sites that offer a "mass submit" service:

Internet Promotions
http://www.websitepromote.com
Gethits.com
http://www.gethits.com
! Register-It!
http://www.register-it.com
www.SitePromoter
http://www.sitepromoter.com
Site See Submission Service
http://www.site-see.com
Submit It!
http://www.submit-it.com
Submit Now
http://www.submitnow.com
WebPromote.com
http://www.webpromote.com

Subject Area Knowledge

Another option for publicizing a Web site is to become familiar with your subject area on the Web. This knowledge has impact at several levels. Subject area knowledge should influence the keywords used in both the metatag information as well as in the top-level content of your site. In the field of health care, Web pages whose audiences include both health-care professionals and health-care consumers should include both medical terms and, where applicable, the layperson equivalent. For example, a Web page with content on myocardial infarction should include both keywords "myocardial infarction" and "heart attack." A library Web page on document delivery should use both library terminology-"interlibrary loan" and "document delivery"-and also language for the layperson-"borrowing."

Subject area knowledge also determines how and with whom you index your site. Find a guide on the Web specific to your subject area or find individual Web sites with similar subject coverage and ask that the Webmaster set up a reciprocal link-"I'll link to you if you link to me." Identify newsgroups or listservs to announce your library's specific services or collections.

Making Web Rings

Reciprocal linking between sites sharing a common interest helps facilitate the creation of user communities and is the impetus behind the creation of Web rings. Web rings are groups of sites sharing a common interest and are offered as an alternative to search engines for locating topic-specific information. According to Webring, Inc., "Member Web sites have banded together to form their sites into linked circles. Their purpose: to allow more visitors to reach them quickly and easily. Through navigation links found most often at the bottom of member pages, visitors can travel all or any of the sites in a ring." Web rings offer a free alternative for publicizing a Web site along with other similar sites-for example, a public library ring. Anyone can become a "ringmaster" by contacting http://www.web ring.com and suggesting a topic-specific Web ring.

More Ways to Keep Your Users Coming Back

Indexing your Web site with the myriad of search sites represents only a small fraction of what can be done to publicize a Web site. There are many more options that with a little creativity offer a more targeted approach to getting the word out. This applies to announcing the initial launch of your site, but also to announcing new additions to the site or updated content. The following is a list of ideas that we use successfully at the University of Virginia Health Sciences Library:

Request that your Web site address be listed with any article about the library or its departments that appears in publications in house and in the community.

Incorporate your Web address into all outgoing paper materials, e.g., letterhead, business cards, brochures, etc. Also include your Web address as part of your e-mail signature file.

Introduce your Web site as part of an open house or conference. Provide a booth with workstations or a kiosk for those who want to browse your site.

Develop and maintain a special e-mail address book of users interested in content changes that occur on your site. Periodically e-mail this group with special announcements of interest, e.g., new class announcements, newly added Web resources, etc.

Request reciprocal links from other departments within your organization or from Webmasters outside your environment who maintain Web pages of interest to your users. Consider starting a Web ring made up of libraries with similar interests, e.g., health.

Keep the page design consistent incorporating sensible navigational elements, along with information zones reflecting page author, modification date, and complete contact information. Keep the site content current and provide a means for users to comment about the site.

Summing It All Up

Publicizing a library Web site is an ongoing activity closely tied to the library's overall public relations and marketing efforts. Assuming the library's audience is already defined, the next steps require developing publicity objectives, determining publicity activities and budget, and assessing how to measure the effectiveness of these efforts. PR activities should integrate tightly with the overall publicity objective. For many of us, the objective is simply to educate user audiences about the resources and services available from the library's Web site.

Closing the PR loop requires ongoing assessment of these efforts. Ask Web site users how they heard about your Web site; pay close attention to the statistics you already collect for possible spikes, e.g., document delivery; and remember to read your Web site's log files to determine if specific PR efforts are paying off. Assess your behind-the-scenes publicity efforts by searching for your Web site across several different search sites. Is your site found with appropriate keywords or word phrases? Is it suitably ranked with a proper description? If not, take a step back: Re-evaluate how your Web page is constructed and review the sites that are listed as more relevant than yours in a search return. Re-assess how the search site indexes its resources and determines relevancy.

Your Web site publicity campaign has two faces: one occurring behind the scenes during the development of your library's Web site content and code, and the other reflecting the public exchange of information taking place between you and your user audience (such as news releases, open houses, etc.). Developing a plan to deliver both sides of this strategy has the greatest potential to efficiently integrate the library Web site into existing business operations and effectively serve the library's users.

Author Affiliation:

Kim Guenther holds a masters in library science from the University of Maryland and has over 7 years of experience developing and managing large-scale Web sites for both nonprofit and for-profit organizations. As the Internet/clinical services coordinator for the University of Virginia Health Sciences Library, she oversees the development of the University Health System Web site, leading an interdisciplinary team providing support to over 150 Health System departmental Web subsites. She also manages the development of the library's Internet and intranet Web sites and related clinically based Web projects serving users throughout the UVA Health System. Her e-mail address is guenther@virginia.edu.

DESCRIPTORS: Web sites; Product introduction; Publicity
SPECIAL FEATURES: Photograph

```
 24/5/10     (Item 10 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2003  Thomson Derwent. All rts. reserv.

010593297    **Image available**
WPI Acc No: 1996-090250/199610
XRPX Acc No: N96-075542
  Information retrieval appts., e.g. electronic notebook connected to
  network - uses  duplicate   deletion  part which leaves only one
  reference result and afterwards deletes other reference result with same
  contents
Patent Assignee: SHARP KK (SHAF  )
Number of Countries: 001  Number of Patents: 001
Patent Family:
Patent No      Kind   Date    Applicat No    Kind   Date     Week
JP 7295994      A   19951110  JP 9485071      A   19940422  199610  B

Priority Applications (No Type Date): JP 9485071 A 19940422
Patent Details:
Patent No  Kind Lan Pg   Main IPC    Filing Notes
JP 7295994    A     16 G06F-017/30

Abstract (Basic): JP 7295994 A
        The appts. is comprised of an input part (1), a data memory (3), a
    reference part (4) and a reference result buffer (5). It also has a
    display part (7) which carries out the displaying of a searched data
    and a reference result.
        A  duplicate   deletion  part (6) is provided in order to leave
    only one reference result after carrying out the deletion process of
    the other reference results which has the same contents with the one
    that has been left to stay.
        ADVANTAGE - Prevents output of reference result having same content
    due.
        Dwg.1/22
Title Terms: INFORMATION; RETRIEVAL; APPARATUS; ELECTRONIC; CONNECT;
  NETWORK; DUPLICATE; DELETE; PART; LEAF; ONE; REFERENCE; RESULT; AFTER;
  DELETE; REFERENCE; RESULT; CONTENT
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI
```

```
 24/5/4     (Item 4 from File: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2003  Thomson Derwent. All rts. reserv.

014036080     **Image available**
WPI Acc No: 2001-520293/200157
XRPX Acc No: N01-385275
  Automatic test and modification for searchable knowledge base, involves
  identifying and  removing  holes and  duplicate  responses and providing
  facilities to modify existing items/adding new items based on response to
  query
Patent Assignee: INTEL CORP (ITLC  )
Inventor: COHEN P M; GALLAGHER W J; KENNEDY R E; SMITH C D; SWANSON G D
Number of Countries: 001  Number of Patents: 001
Patent Family:
Patent No     Kind   Date     Applicat No     Kind   Date      Week
US 6269364     B1  20010731  US 98161163      A   19980925  200157  B

Priority Applications (No Type Date): US 98161163 A 19980925
Patent Details:
Patent No  Kind Lan Pg   Main IPC     Filing Notes
US 6269364    B1    12 G06F-017/30

Abstract (Basic): US 6269364 B1
        NOVELTY - A query is submitted to a searchable knowledge base. The
    hole is identified if no response is received. On receiving responses,
    invalid and duplicate responses are determined. Based on certain
    criteria, duplicate responses and their corresponding entries in the
    database are eliminated automatically. Provisions are given to add new
    items or to modify an existing item of information for providing
    response to some query in future.
        DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the
    following:
        (a) Machine readable storage medium storing machine readable
    instructions;
        (b) Database testing unit
        USE - For testing searchable knowledge base.
        ADVANTAGE - Eliminates duplicate information, identifies mission
    information and verifies the validity of information.
        DESCRIPTION OF DRAWING(S) - The figures show the schema of the
    searchable knowledge base with automatic testing.
        pp; 12 DwgNo 1, 2, 3/6
Title Terms: AUTOMATIC; TEST; MODIFIED; SEARCH; BASE; IDENTIFY; REMOVE;
  HOLE; DUPLICATE; RESPOND; FACILITY; MODIFIED; EXIST; ITEM; ADD; NEW; ITEM
  ; BASED; RESPOND; QUERY
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI
```

```
 24/5/8     (Item 8 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2003  Thomson Derwent. All rts. reserv.

012140057     **Image available**
WPI Acc No: 1998-556969/199847
XRPX Acc No: N98-434223
  Automatic cluster hierarchy generation from large number of documents -
  generating set of unique tokens from documents, with each document
  modelled in cluster with one or more tokens, and with features extracted
  from cluster documents for clustering using features, for subdivision
  into further clusters
Patent Assignee: DIGITAL EQUIP CORP (DIGI  )
Inventor: PRAKASH M; TRAVIS R; VAITHYANATHAN S
Number of Countries: 001  Number of Patents: 001
Patent Family:
Patent No     Kind    Date    Applicat No    Kind    Date     Week
US 5819258     A   19981006  US 97847734     A   19970307  199847  B

Priority Applications (No Type Date): US 97847734 A 19970307
Patent Details:
Patent No  Kind Lan Pg   Main IPC     Filing Notes
US 5819258    A     15 G06F-017/30

Abstract (Basic): US 5819258 A
        The method involves generating a set of unique tokens from the
    documents. Each document is modelled in a cluster with one or more of
    the tokens. Features are extracted from the modelled documents in the
    cluster, and the documents are clustered using the extracted features
    so that the documents in the cluster are subdivided into further
    clusters.
         The process is repeated for each cluster finally generated, until
    a predetermined limit is reached. The generation of unique tokens
    preferably includes separating each document into tokens with a
    predetermined number of delimiters, to generate a pool of tokens, and
    removing  duplicates  from the pool. The pool of tokens is
    pre-processed to eliminate selected tokens which do not represent
    meaningful data.
        USE - For indexing large numbers of documents.
        ADVANTAGE - Organises large sets of documents in response to user
    query, in time efficient and robust manner.
        Dwg.2/6
Title Terms: AUTOMATIC; CLUSTER; HIERARCHY; GENERATE; NUMBER; DOCUMENT;
  GENERATE; SET; UNIQUE; TOKEN; DOCUMENT; DOCUMENT; MODEL; CLUSTER; ONE;
  MORE; TOKEN; FEATURE; EXTRACT; CLUSTER; DOCUMENT; FEATURE; SUBDIVIDED;
  CLUSTER
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI
```

```
 22/5/2     (Item 1 from File: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2003  Thomson Derwent. All rts. reserv.

014281488     **Image available**
WPI Acc No: 2002-102189/200214
XRPX Acc No: N02-075992
  Information search method in Internet, involves deleting redundantly
  registered search  result  using  filter  and displaying search result on
  personal computer
Patent Assignee: NIPPON SYSTEM KIKAKU KK (NISY-N)
Number of Countries: 001  Number of Patents: 001
Patent Family:
Patent No      Kind    Date     Applicat No     Kind    Date      Week
JP 2001344240   A   20011214  JP 2000162271    A   20000531  200214  B

Priority Applications (No Type Date): JP 2000162271 A 20000531
Patent Details:
Patent No  Kind Lan Pg   Main IPC     Filing Notes
JP 2001344240 A      3 G06F-017/30

Abstract (Basic): JP 2001344240 A
        NOVELTY - A  search   engine  file server (1) searches a keyword
    designated from a personal computer (5) through Internet. A data
    registration file (3) registers data searched by several  search
    engines  (11 - 1N). A filter (4) deletes the redundantly registered
    data and finally displays the registered data.
        DETAILED DESCRIPTION - An INDEPENDENT CLAIM is also included for
    search   system  in Internet.
        USE - For searching information in Internet.
        ADVANTAGE - Enables searching data corresponding to desired keyword
    easily and quickly.
        DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of
    the Internet  search   system . (Drawing includes non-English language
    text).
         Search   engine  (11 - 1N)
         Search   engine  file server (1)
        Data registration file (3)
        Filter (4)
        Personal computer (5)
        pp; 3 DwgNo 1/1
Title Terms: INFORMATION; SEARCH; METHOD; DELETE; REGISTER; SEARCH; RESULT;
  FILTER; DISPLAY; SEARCH; RESULT; PERSON; COMPUTER
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI
```

```
 16/5/2     (Item 1 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2003  Thomson Derwent. All rts. reserv.

015227576    **Image available**
WPI Acc No: 2003-288489/200328
XRPX Acc No: N03-229320
   Matching  information merge and data trees prune method for world wide
  web, involves applying merge document to identified  matching  documents
  within source documents
Patent Assignee: INT BUSINESS MACHINES CORP (IBMC  )
Inventor: MYLLYMAKI J P
Number of Countries: 001  Number of Patents: 001
Patent Family:
Patent No     Kind   Date    Applicat No    Kind   Date     Week
US 20020188598  A1  20021212  US 2001834965   A   20010412  200328  B

Priority Applications (No Type Date): US 2001834965 A 20010412
Patent Details:
Patent No  Kind Lan Pg   Main IPC    Filing Notes
US 20020188598 A1    15 G06F-017/30

Abstract (Basic): US 20020188598 A1
        NOVELTY -  Two  or more  source  documents that share a  similar
    data structure are identified.  Matching  documents that relate to the
    same  configurable entity within the  two  or more  source  documents
    are identified. A merge document is applied to the  matching  documents
    to merge the  matching  documents into a  resultant  document, and to
    prune the data tree of the  resultant  document.
        DETAILED DESCRIPTION - An INDEPENDENT CLAIM is included for
    software program product for merging  matching  information and pruning
    data trees.
        USE - For world wide web (WWW).
        ADVANTAGE - Provides a way of retrieving  sets  of individual web
    pages from web sites and locally merging the data. Enables the user to
    obtain logical tree data structure where redundancies have been
    removed . Enables the user to bypass the built-in restrictions in
    product databases to effectively mine the data for information. Permits
     comparative  analysis of the data.
        DESCRIPTION OF DRAWING(S) - The figure shows a schematic view of an
    operation environment utilizing data trees automated merging and
    pruning system.
        pp; 15 DwgNo 1/8
Title Terms:  MATCH ; INFORMATION; MERGE; DATA; TREE; PRUNE; METHOD; WORLD;
  WIDE; WEB; APPLY; MERGE; DOCUMENT; IDENTIFY;  MATCH ; DOCUMENT; SOURCE;
  DOCUMENT
Derwent Class: T01
International Patent Class (Main):  G06F-017/30
International Patent Class (Additional):  G06F-017/24
File Segment: EPI
```

```
Set     Items   Description
S1       3248   SEARCH()(ENGINE? OR SYSTEM?) OR SEARCHENGINE?
S2         95   S1(3N)(MULTIPL? OR MANY OR SEVERAL OR VARIOUS OR PLURAL?) -
             OR METASEARCH? OR METACRAWLER? OR DOGPILE?
S3    1130856   MAP OR MAPPING OR MAPPED OR MAPS OR COMPAR? OR MATCH?
S4    2171954   SIMILAR? OR EQUAL? OR EQUIVALENT? OR SAME?
S5    1487775   REMOV? OR DELET? OR CANCEL?
S6          0   S2 AND S3 AND S4 AND S5
S7         14   S1 AND S4 AND S5
S8        875   DUPLICAT?(2N)(DETECT? OR REMOV? OR DELET?)
S9          0   S2 AND S8
S10         2   S1 AND S8
S11         1   S10 NOT S7
S12   1468052   RESULT? OR ANSWER? OR RETRIEVAL? OR SETS
S13      4172   S3 AND S4 AND S5 AND S12
S14     89049   (SERVER? OR SOURCE? OR SEARCH()ENGINE?)(4N)(MULTIPL? OR PL-
             URAL? OR VARIOUS? OR MANY OR SEVERAL? OR DIFFERENT? OR TWO OR
             SECOND OR ADDITIONAL) OR METACRAWLER? OR METASEARCH? OR META(-
             )(CRAWLER? OR SEARCH?)
S15        49   S13 AND S14
S16         7   S15 AND IC=(G06F? OR H04L?)
S17       534   S12 AND S4(N)5
S18         1   S14 AND S17
S19      2624   FILTER(2N)S12
S20         3   S4 AND S19 AND S14
S21         3   S20 NOT S18
S22         2   S19 AND S1
S23      8805   MC=T01-J05B3
S24        10   S23 AND S8
S25       591   S4(N)S5
S26         0   S23 AND S25
S27     16325   (SIMILAR? OR EQUAL? OR SAME? OR DUPLICATE? OR DUPE)(2N)(RE-
             MOV? OR DELETE? OR ERASE? OR DETECT?)
S28        23   S23 AND S27
S29        15   S28 NOT S24
S30        13   S29 NOT AD>20010730
File 347:JAPIO Oct 1976-2003/Apr(Updated 030804)
         (c) 2003 JPO & JAPIO
File 350:Derwent WPIX 1963-2003/UD,UM &UP=200353
         (c) 2003  Thomson Derwent
```

```
Set     Items   Description
S1       3248   SEARCH()(ENGINE? OR SYSTEM?) OR SEARCHENGINE?
S2         95   S1(3N)(MULTIPL? OR MANY OR SEVERAL OR VARIOUS OR PLURAL?) -
             OR METASEARCH? OR METACRAWLER? OR DOGPILE?
S3    1130856   MAP OR MAPPING OR MAPPED OR MAPS OR COMPAR? OR MATCH?
S4    2171954   SIMILAR? OR EQUAL? OR EQUIVALENT? OR SAME?
S5    1487775   REMOV? OR DELET? OR CANCEL?
S6          0   S2 AND S3 AND S4 AND S5
S7         14   S1 AND S4 AND S5
S8        875   DUPLICAT?(2N)(DETECT? OR REMOV? OR DELET?)
S9          0   S2 AND S8
S10         2   S1 AND S8
S11         1   S10 NOT S7
S12   1468052   RESULT? OR ANSWER? OR RETRIEVAL? OR SETS
S13      4172   S3 AND S4 AND S5 AND S12
S14     89049   (SERVER? OR SOURCE? OR SEARCH()ENGINE?)(4N)(MULTIPL? OR PL-
             URAL? OR VARIOUS? OR MANY OR SEVERAL? OR DIFFERENT? OR TWO OR
             SECOND OR ADDITIONAL) OR METACRAWLER? OR METASEARCH? OR META(-
             )(CRAWLER? OR SEARCH?)
S15        49   S13 AND S14
S16         7   S15 AND IC=(G06F? OR H04L?)
S17       534   S12 AND S4(N)5
S18         1   S14 AND S17
S19      2624   FILTER(2N)S12
S20         3   S4 AND S19 AND S14
S21         3   S20 NOT S18
S22         2   S19 AND S1
S23      8805   MC=T01-J05B3
S24        10   S23 AND S8
S25       591   S4(N)S5
S26         0   S23 AND S25
S27     16325   (SIMILAR? OR EQUAL? OR SAME? OR DUPLICATE? OR DUPE)(2N)(RE-
             MOV? OR DELETE? OR ERASE? OR DETECT?)
S28        23   S23 AND S27
S29        15   S28 NOT S24
S30        13   S29 NOT AD>20010730
File 347:JAPIO Oct 1976-2003/Apr(Updated 030804)
         (c) 2003 JPO & JAPIO
File 350:Derwent WPIX 1963-2003/UD,UM &UP=200353
         (c) 2003  Thomson Derwent
```

24/5/7 (Item 7 from file: 350)
DIALOG(R)File 350:Derwent WPIX
(c) 2003 Thomson Derwent. All rts. reserv.

013455720 **Image available**
WPI Acc No: 2000-627663/200060
XRPX Acc No: N00-465008

Duplicate **query result** removal **program for computerized record keeping system, has instruction corresponding to issue of query, for processing result from each data source before receiving results from next source**

Patent Assignee: NETSCAPE COMMUNICATIONS CORP (NETS-N)
Inventor: GUHA R V
Number of Countries: 001 Number of Patents: 001
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
US 6081805	A	20000627	US 97929352	A	19970910	200060	B

Priority Applications (No Type Date): US 97929352 A 19970910
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 6081805	A		9	G06F-017/30	

Abstract (Basic): US 6081805 A

NOVELTY - The program includes instruction corresponding to issue of query to one data source at a time such that results from each data source is processed before results from next data source is received. Processed results from different data sources are compared to find if there is any matching data based on which duplicated results are discarded.

DETAILED DESCRIPTION - The processing of result obtained from data source is done using Knuth Hashing Algorithms and hash index is prepared. The index of results obtained from two data source is compared to find duplication of data. INDEPENDENT CLAIMS are also included for the following:

(a) **duplicate** query result **removing** process;
(b) **duplicate** query result **removing** apparatus

USE - For **removing duplicate** query result in computerized record keeping system in Internet environment.

ADVANTAGE - Since the result obtained from data sources are compared and duplicate results are discarded before storing in memory, the process of storing all the results in memory and then checking is avoided, hence memory requirement is decreased and also speed in which results are given to user is increased.

DESCRIPTION OF DRAWING(S) - The figure shows flowchart of program of memory duplicate query result.

pp; 9 DwgNo 3/3

Title Terms: DUPLICATE; QUERY; RESULT; REMOVE; PROGRAM; RECORD; KEEP; SYSTEM; INSTRUCTION; CORRESPOND; ISSUE; QUERY; PROCESS; RESULT; DATA; SOURCE; RECEIVE; RESULT; SOURCE
Derwent Class: T01
International Patent Class (Main): G06F-017/30
File Segment: EPI